DP-302096 (DEL01 P-333)

SPEECH RECOGNITION WITH USER SPECIFIC ADAPTIVE VOICE FEEDBACK

This application is a continuation-in-part of U.S. Patent Application Serial No. TBD (Docket No. DP-301962) filed January 10, 2000, entitled "SPEECH RECOGNITION WITH ADAPTIVE VOICE FEEDBACK" to Scott A. Deyoe, Tuan A. Hoang and Shishong Huang.

Background of the Invention

The present invention is directed to speech recognition, and more specifically to providing user specific adaptive voice feedback in a multi-level speech recognition driven system.

As is well known to one of ordinary skill in the art, speech recognition is a field in computer science that deals with designing computer systems that can recognize spoken words. A number of speech recognition systems are currently available (e.g., products are offered by IBM, Dragon Systems, Lernout & Hauspie and Philips). Traditionally, speech recognition systems have only been used in a few specialized situations due to their cost and limited functionality. For example, such systems have been implemented when a user was unable to use a keyboard to enter data because the user's hands were disabled. Instead of typing commands, the user spoke into a microphone. However, as the cost of these systems has continued to decrease and the performance of these systems has continued to increase, speech recognition systems are being used in a wider variety of applications (as an alternative to keyboards or other user interfaces). For example, speech actuated control systems have been implemented in motor vehicles to control various accessories within the motor vehicles.

A typical speech recognition system, that is implemented in a motor vehicle, includes voice processing circuitry and memory for storing data representing command words (that are employed to control various vehicle accessories). In a typical system, a microprocessor is utilized to

compare the user provided data (i.e., voice input) to stored speech models to determine if a word match has occurred and provide a corresponding control output signal in such an event. The microprocessor has also normally controlled a plurality of motor vehicle accessories, e.g., a cellular telephone and a radio. Such systems have advantageously allowed a driver of the motor vehicle to maintain vigilance while driving the vehicle.

Some speech recognition systems also recognize (by utilizing voice recognition technology) a specific user. However, most current speech recognition systems require a user to learn unique wording and dialogs for successful operation of the system. Many of these systems have very long voice dialog prompts to direct a user such that the dialog can progress. Further, the help function of most of these systems has required the user to request assistance via a voice command, such as "Help" or "What can I say?" at which point the user is then provided with an available word or dialog option. These systems have typically been inflexible and not readily adaptable as the ability of the user of the system changed.

As such, a speech recognition system that adapts to a specific user by providing assistance automatically and only as needed is desirable.

Summary of the Invention

The present invention is directed to a technique for providing user specific adaptive voice feedback in a multi-level speech recognition driven system. Initially, the system detects whether a user of the system has provided a voice input. If a voice input is detected, the system then determines whether the voice input is associated with a specific user that is recognized by the system. If the user has not provided a voice input for a predetermined user specific time period, the system provides adaptive voice feedback to the user. When the system receives or detects a voice input from the user, the system determines whether the input is recognized. If the input is recognized by the system, the speech selectable task that corresponds to the input is performed. In another embodiment, when the user has failed to

respond for a user specific set number of the predetermined user specific time periods, at a given level, the system is deactivated.

These and other features, advantages and objects of the present invention will be further understood and appreciated by those skilled in the art by reference to the following specification, claims and appended drawings.

Brief Description of the Drawings

The present invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Fig. 1 is a block diagram of a speech recognition system used in a motor vehicle;

Figs. 2A-2C are a flow diagram of an adaptive voice feedback routine, according to an embodiment of the present invention;

Fig. 3 is an exemplary dialog tree that can be implemented with an adaptive voice feedback system, according to an embodiment of the present invention; and

Figs. 4A-4C are a flow diagram of an adaptive voice feedback routine, according to another embodiment of the present invention.

Description of the Preferred Embodiments

Fig. 1 is a block diagram of a speech recognition system 100 (implemented within a motor vehicle) that provides adaptive voice feedback, according to an embodiment of the present invention. System 100 includes a processor 102 coupled to a motor vehicle accessory 124 and a display 120. Processor 102 controls motor vehicle accessory 124, at least in part, as dictated by voice input supplied by a user of system 100. Processor 102 also supplies various information to display 120, to allow a user of the motor vehicle to better utilize system 100. In this context, the term processor may include a general purpose processor, a microcontroller (i.e., an execution unit with memory, etc., integrated within a single integrated circuit) or a digital signal processor (DSP). Processor 102 is also coupled to a memory

subsystem 104. Memory subsystem 104 includes an application appropriate amount of main memory (volatile and non-volatile).

An audio input device 118 (e.g., a microphone) is coupled to a filter/amplifier module 116. Filter/amplifier module 116 filters and amplifies the voice input provided by the user through audio input device 118. Filter amplifier module 116 is also coupled to an analog-to-digital (A/D) converter 114. A/D converter 114 digitizes the voice input from the user and supplies the digitized voice to processor 102 (which causes the voice input to be compared to system recognized commands).

Processor 102 executes various routines in determining whether the voice input corresponds to a system recognized command. Processor 102 also causes an appropriate voice output to be provided to the user (ultimately through an audio output device 112). The synthesized voice output is provided by the processor 102 to a digital-to-analog (D/A) converter 108. D/A converter 108 is coupled to a filter/amplifier section 110, which amplifies and filters the analog voice output. The amplified and filtered voice output is then provided to audio output device 112 (e.g., a speaker). While only one motor vehicle accessory module 124 is shown, it is contemplated that any number of accessories, typically provided in a motor vehicle (e.g., a cellular telephone or a radio), can be implemented.

Processor 102 may execute a routine or may be coupled to an adaptable module 126 (which can include artificial intelligence (AI) code, fuzzy logic, a neural network or any other such appropriate technology) that can identify the specific dialogs a specific user has mastered and those dialogs that require additional assistance. This enables the system to adjust the timing in which assistance, in the form of adaptive voice feedback, is provided to a specific user and is further discussed in conjunction with Figs. 4A-4C.

Figs. 2A-2C are a flow diagram of an adaptive voice feedback routine 200, according to an embodiment of the present invention. In the embodiment of Figs. 2A-2C, routine 200 determines whether a voice input, provided by a user, corresponds to, for example, a command. Routine 200 does not identify specific users. In step 202, the multi-level speech

recognition driven system is activated. A user may activate the system, for example, through a voice command or by physically asserting a switch. From step 202, control transfers to step 204. In step 204, a first level variable "pass1", which tracks the number of times that a first level idle timer has expired, is initialized. From step 204, control transfers to step 206. In step 206, routine 200 determines whether a voice input has been detected. If so, control transfers from step 206 to step 216. If not, control transfers from step 206 to step 208. In step 208, routine 200 determines whether a first level idle timer has expired. If the first level idle timer has not expired in step 208, control transfers to step 206. If the first level idle timer has expired, control transfers from step 208 to step 210. In step 210, routine 200 causes the "pass1" variable to be incremented and resets the first level idle timer.

From step 210, control transfers to step 212. In step 212, routine 200 determines whether the "pass1" variable has exceeded a set value (in this case, three). One of ordinary skill in the art will appreciate that the decision threshold for the "pass1" variable can be adjusted, as desired. If the "pass1" variable is less than or equal to three, control transfers from step 212 to step 214. In step 214, routine 200 provides a first level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 214, control transfers to step 206. In step 212 if the "pass1" variable has exceeded the set value, control transfers to step 248. In step 248, routine 200 causes the speech recognition system to be deactivated. From step 248, control transfers to step 250 where the routine 200 ends.

In step 206, if voice input is detected, control transfers to step 216. In step 216, routine 200 determines whether the voice input is a recognized command. If so, control transfers from step 216 to step 218 (see Fig. 2B). If not, control transfers from step 216 to step 248. One of ordinary skill in the art will appreciate that if the voice input is not recognized, control can alternatively be transferred to step 214 (where first level adaptive voice feedback is provided). In step 218, a second level variable "pass2", which tracks the number of times that a second level idle timer has expired, is initialized. From step 218, control transfers to step 220. In step 220, if voice

input is not detected, control transfers to step 224. In step 224, routine 200 determines whether the second level idle timer has expired. If the second level idle timer has not expired, control transfers from step 224 to step 220. If the second level idle timer has expired, control transfers from step 224 to step 226.

In step 226, routine 200 causes the "pass2" variable to be incremented and resets the second level idle timer. In step 228, routine 200 determines whether the "pass2" variable has exceeded a set value (in this case, three). One of ordinary skill in the art will appreciate that the decision threshold for the "pass2" variable can be adjusted, as desired. If the "pass2" variable has exceeded the set value, control transfers from step 228 to step 248. If the "pass2" variable has not exceeded the set value, control transfers from step 228 to step 230. In step 230, routine 200 provides an appropriate second level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 230, control transfers to step 220. In step 220, if voice input is detected, control transfers to step 232. In step 232, routine 200 determines whether the voice input is recognized. If so, control transfers from step 232 to step 234 (see Fig. 2C). If not, control transfers from step 232 to step 248. One of ordinary skill in the art will appreciate that, if the voice input is not recognized, control can alternatively be transferred to step 230 (where second level adaptive voice feedback is provided).

In step 234, a third level variable "pass3", which tracks the number of times that a third level idle timer has expired, is initialized. From step 234, control transfers to step 236. In step 236, routine 200 determines whether voice input is detected. If so, control transfers to step 246. If not, control transfers from step 236 to step 238. In step 238, routine 200 determines whether the third level idle timer has expired. If the third level idle timer has not expired in step 238, control transfers to step 236. If the third level idle timer has expired, control transfers from step 238 to step 240. In step 240, routine 200 causes the "pass3" variable to be incremented and resets the third level idle timer.

From step 240, control transfers to step 242. In step 242, routine 200 determines whether the "pass3" variable has exceeded a set value (in this case, three). One of ordinary skill in the art will appreciate that the decision threshold for the "pass3" (as well as "pass1" and "pass2") variable can be adjusted, as desired. If the "pass3" variable is less than or equal to three, control transfers from step 242 to step 244. In step 244, routine 200 provides an appropriate third level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 244, control transfers to step 236. In step 242 if the "pass3" variable has exceeded the set value, control transfers to step 248. In step 248, routine 200 causes the speech recognition system to be deactivated. From step 248, control transfers to step 250 where the routine 200 ends.

In step 246, routine 200 determines whether the voice input is recognized. If so, control transfers from step 246 to step 252. If not, control transfers from step 246 to step 248. One of ordinary skill in the art will appreciate that, if the voice input is not recognized control can alternatively be transferred to step 244 (where third level adaptive voice feedback is provided). In step 252, routine 200 causes the voice selected task to be ran. From step 252, control transfers to step 250 where routine 200 ends. Thus, a system has been described, which provides adaptive voice feedback when appropriate. This can be determined at each level by setting a level dependent idle timer to a particular value. Alternatively, the idle timer can be dialog branch dependent. As mentioned above, the number of times in which the idle timer is allowed to expire at a given level is also adjustable. As such, a system according to the present invention provides adaptive voice feedback that is appropriate for the experience level of the user. For example, if a user is inexperienced, the system will provide voice feedback at each level. However, if a user is experienced, the user can provide continuous voice input to the system and the system will not provide voice feedback to the user.

This allows a novice user to begin immediately using the speech recognition system without having to first study a user's guide. By

monitoring the time since a voice input was last received (to determine whether to activate the adaptive voice feedback), the system can be advantageously used with a wide range of users with different experience levels. As discussed above, the system provides a context sensitive voice prompt, as required, to continue the voice dialog. A user may wait for adaptive voice feedback to complete the user's selection or the user may 'barge-in' with a desired command or use a word such as 'yes' or 'select' to indicate a desired option. While a three level dialog has been described, one of ordinary skill in the art will readily appreciate that the present invention can be implemented with systems that employ a different number of levels.

Fig. 3 is an exemplary dialog tree 300 that further illustrates the functioning of the adaptive voice feedback feature, according to an embodiment of the present invention. At entry point 302 a user activates the speech driven system by speaking the keyword "start". An experienced user that already knows the functions that the user wants performed can then speak the commands in successive order. For example, an experienced user might speak the command string "A, A1, A1C" or "A, A2". On the other hand, an inexperienced user may hesitate after speaking the keyword "start", at which point the system supplies the first level commands "A, B or C" (corresponding to entry points 304, 306 and 308, respectively), after a predetermined time period. If an inexperienced user speaks the command "A" and then hesitates, the system supplies the second level commands "A1, A2 or A3" (corresponding to entry points 310, 312 and 314, respectively), after a predetermined time period. At that point, if an inexperienced user speaks the command "A1" and then hesitates, the system supplies the third level commands "A1A, A1B, A1C or A1D" (corresponding to entry points 316, 318, 320 and 322, respectively), after a predetermined time period.

Thus, if at any level a user is unsure of the next command, after an appropriate period, the system supplies an appropriate voice feedback with a list of commands necessary to continue. Thus, an inexperienced user can learn the system dialog while using the system. A user may receive a prompt after each spoken command because of the hesitation in thinking of the

next word. On the other hand, an experienced user can immediately say all the words in a command and not receive a prompt. As another example, a driver of a motor vehicle attempting to utilize a radio within the motor vehicle may use the command "radio" to activate the radio. At that point, the driver may, for example, have the option of saying "AM", "FM", "tune", "mute", "balance" or "scan". If the driver provides the voice command "tune", the driver may have the option of tuning "up" or "down". Alternatively, the driver may enter a radio channel using a command string, such as, "Radio, FM, Channel, 101.1."

Figs. 4A-4C are a flow diagram of a user specific adaptive voice feedback routine 400, according to another embodiment of the present invention. In the embodiment of Figs. 4A-4C, routine 400 identifies a specific user from a voice input, as well as determining if the voice input corresponds to a particular system recognized input (e.g., command). In step 402, the multi-level speech recognition driven system is activated. While this example is directed to a speech activated system, one of ordinary skill in the art will appreciate that the techniques described herein can readily be applied to a switch activated system. In a switch activated system, the switch is typically monitored by an input of processor 102. From step 402, control transfers to step 404. In step 404, a first level variable "pass1", which tracks the number of times that a user specific first level idle timer has expired, is initialized. From step 404, control transfers to step 406. In step 406, routine 400 determines whether a specific user is recognized by the speech recognition system. In a speech activated system, the voice input provided by a user, to activate the system, is compared (using commercially available voice recognition technology) to a plurality of established user voice patterns, if any.

The established user voice patterns are utilized in recognizing a specific user. If the specific user is recognized by the system, control transfers from step 406 to step 408. In step 408, a user profile that corresponds to the specific user is selected. The specific user profile establishes a predetermined user specific time period for a given level or

dialog branch. The specific user profile also establishes a maximum loop count (a user specific set number that corresponds to the predetermined user specific time periods that are allowed to expire at a given level or dialog branch, before the system is deactivated).

In one embodiment, the predetermined user specific time period and the maximum loop count are adjusted by the system as the ability of the specific user changes. For example, as a specific user of the system becomes more familiar with the system these values are decreased. One of skill in the art will appreciate that the values can be adjusted, as desired. This can be readily accomplished by utilizing artificial intelligence code, fuzzy logic, neural networks or other such adaptable networks, well known to one of ordinary skill in the art, that track the ability of each user. From step 408, control then transfers to step 412.

If the specific user is not recognized by the system (e.g., a new user) in step 406, control transfers to step 410 where a default user profile is established. Thereafter, a profile for that new user is stored within the system such that when that user utilizes the system again, the profile for that specific user is selected. One of ordinary skill in the art will readily appreciate that the number of such new users that can be added to the system is only limited by the system resources (e.g., volatile and non-volatile memory, processing power, etc.). From step 410, control transfers to step 412. In step 412, routine 400 determines whether a voice input has been detected. One of skill in the art will appreciate that if the system is not voice activated, the determination of the specific user would occur after a voice input (e.g., a spoken command) is received.

If a voice input is detected, control transfers from step 412 to step 422. If not, control transfers from step 412 to step 414. In step 414, routine 400 determines whether a user specific first level idle timer has expired. If the first level idle timer has not expired in step 414, control transfers to step 412. If the first level idle timer has expired, control transfers from step 414 to step 416. In step 416, routine 400 causes the "pass1"

variable to be incremented and resets the first level idle timer. As discussed above, the value of the first level idle timer is user specific.

From step 416, control transfers to step 418. In step 418, routine 400 determines whether the "pass1" variable is less than a maximum loop count (i.e., a user specific set number that indicates the number of times that a predetermined user specific time period has expired). As discussed above, the decision threshold for the "pass1" variable is user specific and is adjusted by the adaptable module 126 or a routine running on processor 102. If the "pass1" variable is less than the maximum loop count, control transfers from step 418 to step 420. In step 420, routine 400 provides a first level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 420, control transfers to step 412. In step 418 if the "pass1" variable has exceeded the maximum loop count, control transfers to step 454. In step 454, routine 400 causes the speech recognition system to be deactivated. From step 454, control transfers to step 456 where routine 400 ends.

In step 412, if voice input is detected, control transfers to step 422. In step 422, routine 400 determines whether the voice input is recognized. If so, control transfers from step 422 to step 424 (see Fig. 4B). If not, control transfers from step 422 to step 420 (where first level adaptive voice feedback is provided). One of ordinary skill in the art will appreciate that if the voice input is not recognized, control can alternatively be transferred to step 454. In step 424, a second level variable "pass2", which tracks the number of times that a user specific second level idle timer has expired, is initialized. From step 424, control transfers to step 426. In step 426, if voice input is not detected, control transfers to step 428. In step 428, routine 400 determines whether the second level idle timer has expired. If the second level idle timer has not expired, control transfers from step 428 to step 426. If the second level idle timer has expired, control transfers from step 428 to step 430.

In step 430, routine 400 causes the "pass2" variable to be incremented and resets the second level idle timer. From step 430, control

then transfers to step 432. In step 432, routine 400 determines whether the "pass2" variable is less than a maximum loop count (i.e., a user specific set number that indicates the number of times that a predetermined user specific time period has expired). As discussed above, the decision threshold for the "pass2" variable is user specific and is adjusted by the system as determined by adaptable module 126 or a routine running on processor 102. If the "pass2" variable is not less than the maximum loop count, control transfers from step 432 to step 454. If the "pass2" variable is less than the maximum loop count, control transfers from step 432 to step 434.

In step 434, routine 400 provides an appropriate second level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 434, control transfers to step 426. In step 426, if voice input is detected, control transfers to step 436. In step 436, routine 400 determines whether the voice input is recognized. If so, control transfers from step 436 to step 438 (see Fig. 4C). If not, control transfers from step 436 to step 434 (where second level adaptive voice feedback is provided). One of ordinary skill in the art will appreciate that, if the voice input is not recognized, control can alternatively be transferred to step 454.

In step 438, a third level variable "pass3", which tracks the number of times that a user specific third level idle timer has expired, is initialized. From step 438, control transfers to step 440. In step 440, routine 400 determines whether voice input is detected. If so, control transfers to step 450. If not, control transfers from step 440 to step 442. In step 442, routine 400 determines whether the third level idle timer has expired. If the third level idle timer has not expired in step 442, control transfers to step 440. If the third level idle timer has expired, control transfers from step 442 to step 444. In step 444, routine 400 causes the "pass3" variable to be incremented and resets the third level idle timer.

From step 444, control transfers to step 446. In step 446, routine 400 determines whether the "pass3" variable has exceeded a maximum loop count. As discussed above, the decision threshold for the

"pass3" (as well as "pass1" and "pass2") variable is adjusted as the experience level of each specific user changes. If the "pass3" variable is less than the maximum loop count, control transfers from step 446 to step 448. In step 448, routine 400 provides an appropriate third level adaptive voice feedback. This allows a user to determine which command should be spoken at that time. From step 448, control transfers to step 440. In step 446 if the "pass3" variable is not less than the maximum loop count, control transfers to step 454. In step 454, routine 400 causes the speech recognition system to be deactivated. From step 454, control transfers to step 456 where the routine 400 ends

In step 450, routine 400 determines whether the voice input is recognized. If so, control transfers from step 450 to step 452. If not, control transfers from step 450 to step 448 (where third level adaptive voice feedback is provided). One of ordinary skill in the art will appreciate that, if the voice input is not recognized, control can alternatively be transferred to step 454. In step 452, routine 400 causes the voice selected task to be ran. From step 452, control transfers to step 456 where routine 400 ends.

Thus, an alternative system has been described, which provides user specific adaptive voice feedback, when appropriate. Adaptable module 126 or a routine running on processor 102 functions to change a user specific idle timer, that is either level or dialog branch dependent. As mentioned above, adaptable module 126 or a routine running on processor 102 also functions to change the number of times in which the user specific idle timer is allowed to expire. This can be level or dialog branch dependent. As such, a system according to the present invention provides adaptive voice feedback that is presented at an appropriate time for a specific user. When a specific user advances in knowledge of the system, the system adjusts the idle timers for that user. In this manner, the time frame in which voice feedback is provided is customized for each recognized user. While a three level dialog has been described, one of ordinary skill in the art will readily appreciate that this embodiment of the present invention can be implemented with systems that employ a different number of levels.

The above description is considered that of the preferred embodiments only. Modification of the invention will occur to those skilled in the art and to those who make or use the invention. Therefore, it is understood that the embodiments shown in the drawings and described above are merely for illustrative purposes and not intended to limit the scope of the invention, which is defined by the following claims as interpreted according to the principles of patent law, including the Doctrine of Equivalents.